**Regression** analysis attempts to predict one dependent variable or target (usually denoted by Y) and a series of other independent variables or features (usually denoted by X).

**What Is Regression?**

Regression searches for relationships among variables.

For example, you can observe several employees of some company and try to understand how their salaries depend on the **features**, such as experience, level of education, role, a city they work in, and so on.

This is a regression problem where data related to each employee represent one **observation**. The presumption is that the experience, education, role, and city are the independent features, while the salary depends on them.

Similarly, you can try to establish a mathematical dependence of the prices of houses on their areas, numbers of bedrooms, distances to the city centre, and so on.

Generally, in regression analysis, you usually consider some phenomenon of interest and have many observations. Each observation has two or more features. Following the assumption that (at least) one of the features depends on the others, you try to establish a relation among them.

In other words, **you need to find a function that maps some features or variables to others sufficiently well**.

The dependent features are called the **dependent variables**, **outputs**, or **responses**.

The independent features are called the **independent variables**, **inputs**, or **predictors**.

Regression problems usually have one continuous and unbounded dependent variable. The inputs, however, can be continuous, discrete, or even categorical data such as gender, nationality, brand, and so on.

It is a common practice to denote the outputs with 𝑦 and inputs with 𝑥. If there are two or more independent variables, they can be represented as the vector 𝐱 = (𝑥₁, …, 𝑥ᵣ), where 𝑟 is the number of inputs.

**When Do You Need Regression?**

Typically, you need regression to answer whether and how some phenomenon influences the other or **how several variables are related**. For example, you can use it to determine *if* and *to what extent* the experience or gender impact salaries.

Regression is also useful when you want **to forecast a response** using a new set of predictors. For example, you could try to predict electricity consumption of a household for the next hour given the outdoor temperature, time of day, and number of residents in that household.

Regression is used in many different fields: economy, computer science, social sciences, and so on. Its importance rises every day with the availability of large amounts of data and increased awareness of the practical value of data.

**Linear Regression**

**Linear regression**is a statistical approach for modelling the relationship between a dependent variable with a set of explanatory variables. Linear regression is a common Statistical Data Analysis technique.

Problem-solving using linear regression has so many applications in business, social, biological, and many other areas.

**Types of Linear Regression**

In general, there are two types of linear regression

* Simple Linear Regression
* Multiple Linear Regression

**Simple Linear Regression** allows us to study the relationship between two variables.

In simple linear regression a single independent variable is used to predict the value of a dependent variable.

I.e. One independent variable (X) and One Dependent variable (Y).

This can be denoted by

**yi = b0 + b1x + e**

In school, we became familiar with equations like the one shown below, then how it is different from the equation above?

**Y = mx + c**

Well both are the same type of equation, we just changed the name of the variable and added something extra, the ‘e’ to minimize the chance of error.

Where **m** was a slope, and c was Intercept. In the first equation, **b0** is the intercept and **b1** is the slope.

**Slope direction**

The slope of a line can be positive, negative, zero, or undefined.

**Positive slope:** y increases as x increases, so the line slopes upwards to the right.

**Negative slope:** y decreases as x increases, so the line slopes downwards to the right. If you remember from the previous examples, we have seen an example of this, where the Age of the Car increases, the price decreases.

**Zero slope:** y does not change as x increases, so the line remains horizontal. The slope of any horizontal line is always zero.

**Undefined slope:** When the line is exactly vertical, it does not have a defined slope. The two x coordinates are the same, so the difference is zero.

**Multiple Linear Regression** allows us to study the relationship between three or more variables.

In Multiple Linear Regression two or more independent variables are used to predict the value of a dependent variable.

I.e. Two or more independent variables (X1, X2, X3, ….) and one Dependent variable (Y).

**yi= b0+ b1x1+ b2x2+ b3x3 + ... + e**

The difference between linear regression and multiple linear regression is the number of independent variables (X). In both cases, there is only a single dependent variable (Y).

There is one more type of Linear Regression which is Polynomial Regression, this is a generalized case of linear regression, which we will come in another paper and, we will find that many of the challenges we encounter, can be solved with Multiple Linear Regression.